The Coyote Universe

Salman Habib, T-2; Katrin Heitmann, ISR-1; David Higdon, Earl Lawrence, CCS-6; Christian Wagner, Astrophysikalisches Institut Potsdam; Martin White, University of California, Berkeley; Brian J. Williams, CCS-6

Over the last decade, a powerful and diverse suite of cosmological observations has led to remarkable discoveries in fundamental physics. Dark energy and dark matter form the nexus of an intense global endeavor to fathom the deep mysteries posed by their existence. Remarkably, the statistical quality of cosmological measurements is such that our ability to interpret them will become theory-limited in the very near future, if the state-of-the-art in modeling and simulation of structure formation in the Universe cannot keep pace with the observations.

In order to overcome this challenge, structure-formation probes of dark energy and dark matter are the current focus area of a computational cosmology program at LANL. This program covers studies of the galaxy and galaxy cluster distribution, cosmological mass mapping via weak gravitational lensing observations, and baryon acoustic oscillations—the signature of acoustic waves in the primordial fireball as imprinted on the distribution of galaxies today—as a probe of the geometry of the Universe.

A major new effort in this area is represented by the Coyote Universe simulation suite, named after the high-performance computing cluster at LANL on which the simulations were carried out. The simulation suite ranges over 38 sets of cosmological parameters and consists of 800 simulations with an associated database of 60 TB—one of the very largest cosmological simulation databases today.

The requisite error constraints on theoretical predictions are very tight—for the temperature anisotropy in the cosmic microwave background, at better than 0.1 %, and only slightly worse for statistical measures of the large-scale distribution of matter, such as the two-point correlation function or, equivalently, the fluctuation power spectrum. These levels of accuracy have rarely been attained for theoretical predictions of complex nonlinear structure formation in any field, not just in cosmology.

The second important problem is that inverse analysis of cosmological observational datasets requires running many simulations with different cosmological parameters, typically tens to hundreds of thousands of simulations. Given the temporal cost of large-scale supercomputer runs, this is far from being achievable in practice. A more realistic target is in the range of tens to perhaps hundreds of simulations.

Consequently, our approach is first to show that it is possible to attain the required subpercent error control in numerical simulations of the formation of structure in the Universe, and following that, to show how only a limited set of high-quality simulations can take the place of tens to hundreds of thousands of simulations. This can be achieved using a statistical methodology developed by us—the Cosmic Calibration Framework [1-3].

The nonlinear epoch of structure formation in the Universe, responsible for the dark matter clumps within which individual galaxies, groups, and clusters of galaxies reside, is studied via N-body simulations. The matter density is represented by a large number of point particles interacting

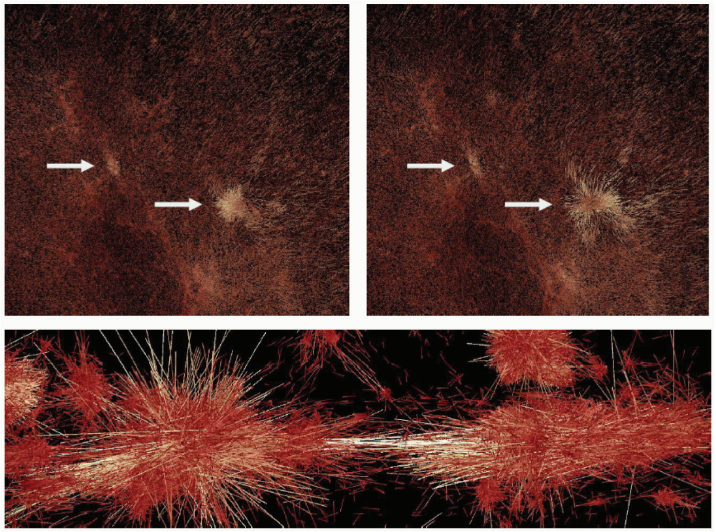

Fig. 1. Top two panels: Early (left) and late (right) start simulations compared. Particle color coding is with respect to velocity. Note the less-tight structures due to the late start (arrows). The bottom panel shows this behavior along a density filament, displayed as differences between the two simulation particle positions (the needle-like lines).

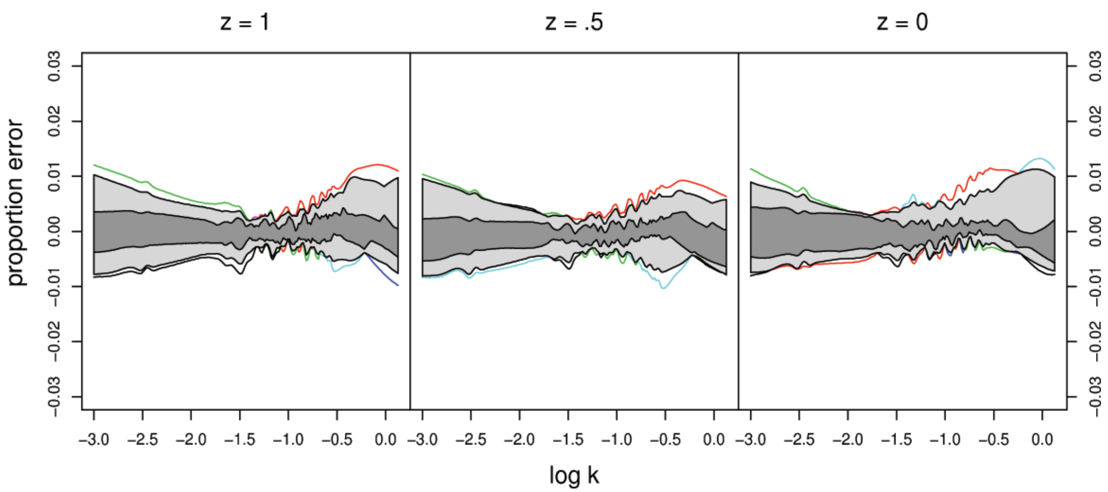

Fig. 2. Emulator performance at three cosmological epochs, with red shifts, z = 1, 0.5, and 0 (left to right). The emulator is tested on 10 additional runs — runs not used to build it — within the parameter priors. The emulator error with respect to these results is shown. The central gray region contains the middle 50 % of the residuals, the wider light gray region, the middle 90 %. Errors are at the subpercent level.

via gravity in an expanding Universe. To control errors in these simulations, we have identified several key factors, some of which had heretofore not been properly understood. One of these factors is the importance of beginning the simulations sufficiently early on in the history of the Universe—start too late and structures tend to be too diffuse (Fig. 1). Special visualization techniques were developed in order to be able to see and understand such subtle effects. Other components of precision simulations, as identified by us, include very large simulation volumes, accurate time-stepping, and sufficient mass resolution.

The reduction in the required number of simulations is made possible by a sophisticated interpolation scheme ("emulation") applicable to high-dimensional datasets. This interpolation scheme uses several of the most powerful ideas in modern Bayesian statistical methods: advanced sampling schemes, Gaussian process modeling, and principal component basis representation of the data. With a sampling scheme based on only 37 simulations, it is possible to accurately predict several observational quantities to subpercent accuracy for any cosmological model with parameters that fall within a preset prior, determined (conservatively) from the current uncertainties in measured cosmological parameters. Figure 2 shows an example of the tight error control possible with our emulation methodology.

Results from the Coyote Universe simulations are being reported in an initial series of three papers: the first showing that sufficiently accurate simulations can in fact be carried out [4], and the second showing that a limited number of simulations can be bootstrapped to represent the results of many more simulations with little or no loss of accuracy [5]. The third and final paper will encapsulate the results from all the simulations in terms of "emulators," very fast interpolators, which are based on the results from the simulations and can give predictions for observable quantities for any cosmology within some prescribed cosmological parameter range [6]. The final dataset will be made publicly available.

For further information contact Salman Habib at habib@lanl.gov.

[1] M. Schneider, et al., *Phys. Rev. D* **78,** 063529 (2008).
[2] S. Habib, et al., *Phys. Rev. D* **76,** 083503 (2007).
[3] K. Heitmann, et al., *Astrophys. J. Lett.* **646,** L1 (2006).
[4] K. Heitmann, et al., *Astrophys. J.* (submitted) arXiv:0812.1052v1.
[5] K. Heitmann, et al., *Astrophys. J.* (to be submitted).
[6] E. Lawrence, et al. (in preparation).

Funding Acknowledgments
- DOE, Office of Science, High Energy Physics, Dark Energy Program
- LANL Directed Research and Development Program
- LANL Institutional Supercomputing Awards
- LANL IGPP Collaboration Awards Program